FIG. 1

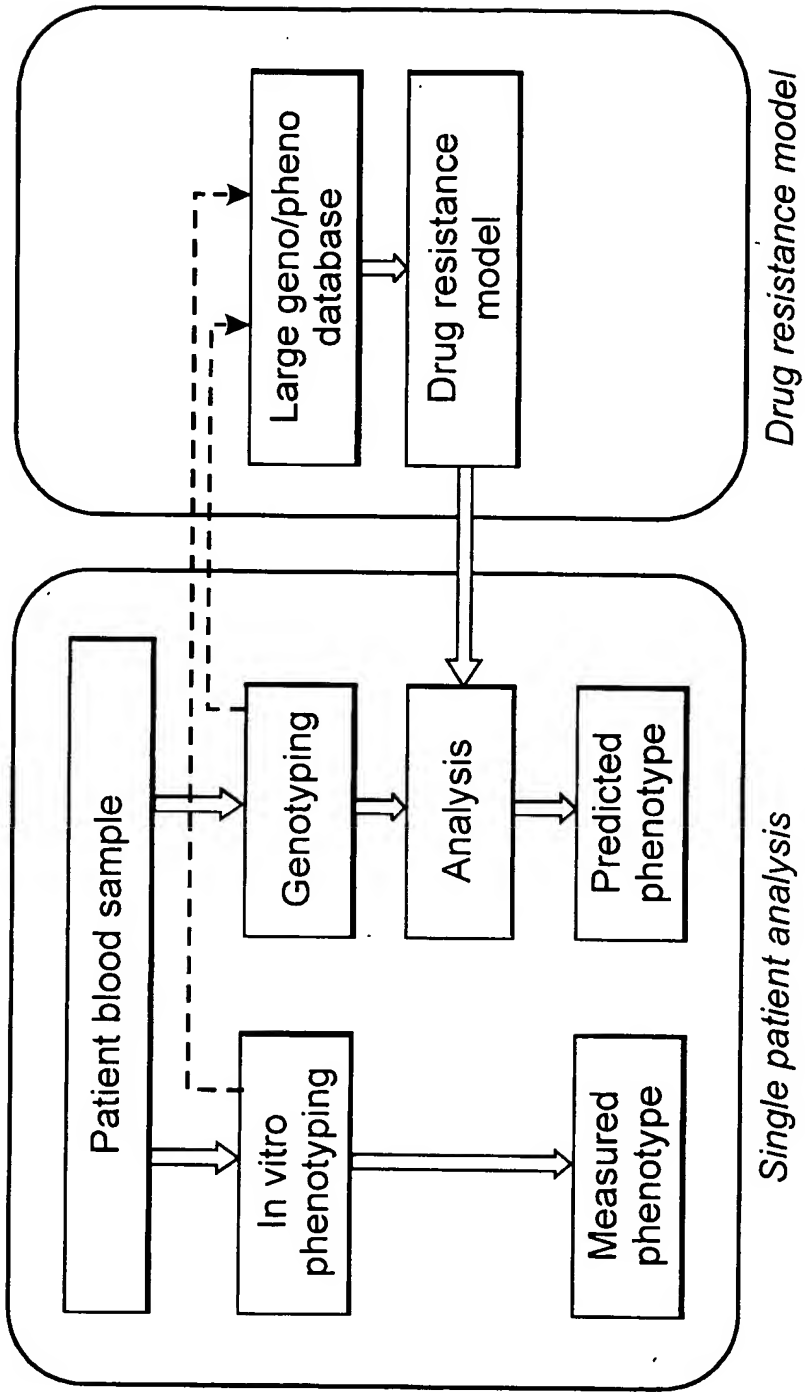

Large geno/pheno database
Drug resistance model
Drug resistance model
Patient blood sample
Genotyping
Analysis
Predicted phenotype
In vitro phenotyping
Measured phenotype
Single patient analysis

FIG. 2

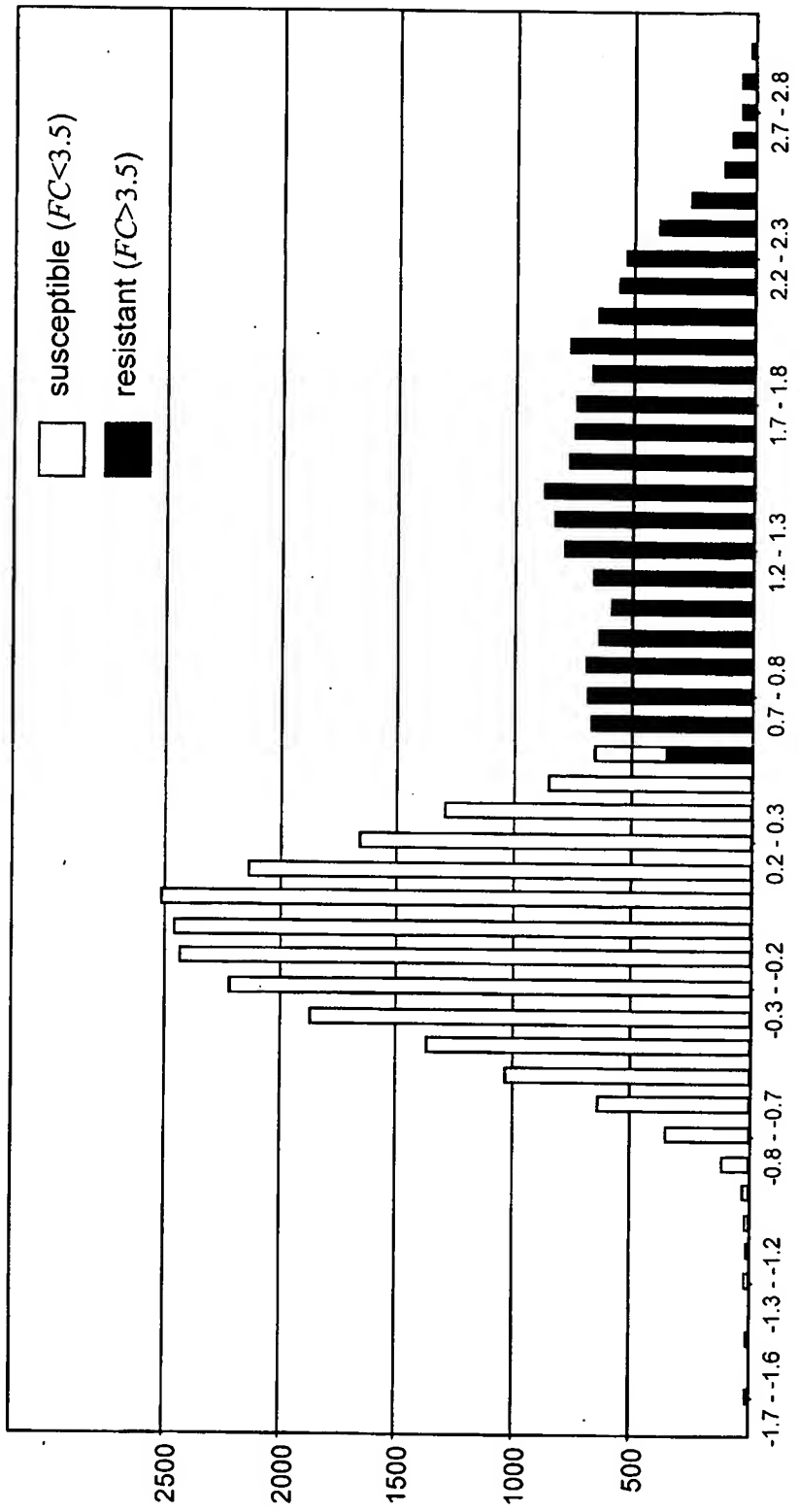

susceptible (FC<3.5)
resistant (FC>3.5)
2500
2000
1500
1000
500
-1.7 - -1.6
-1.3 - -1.2
-0.8 - -0.7
-0.3 - -0.2
0.2 - 0.3
0.7 - 0.8
1.2 - 1.3
1.7 - 1.8
2.2 - 2.3
2.7 - 2.8
RTV log(FC) distribution

FIG. 3

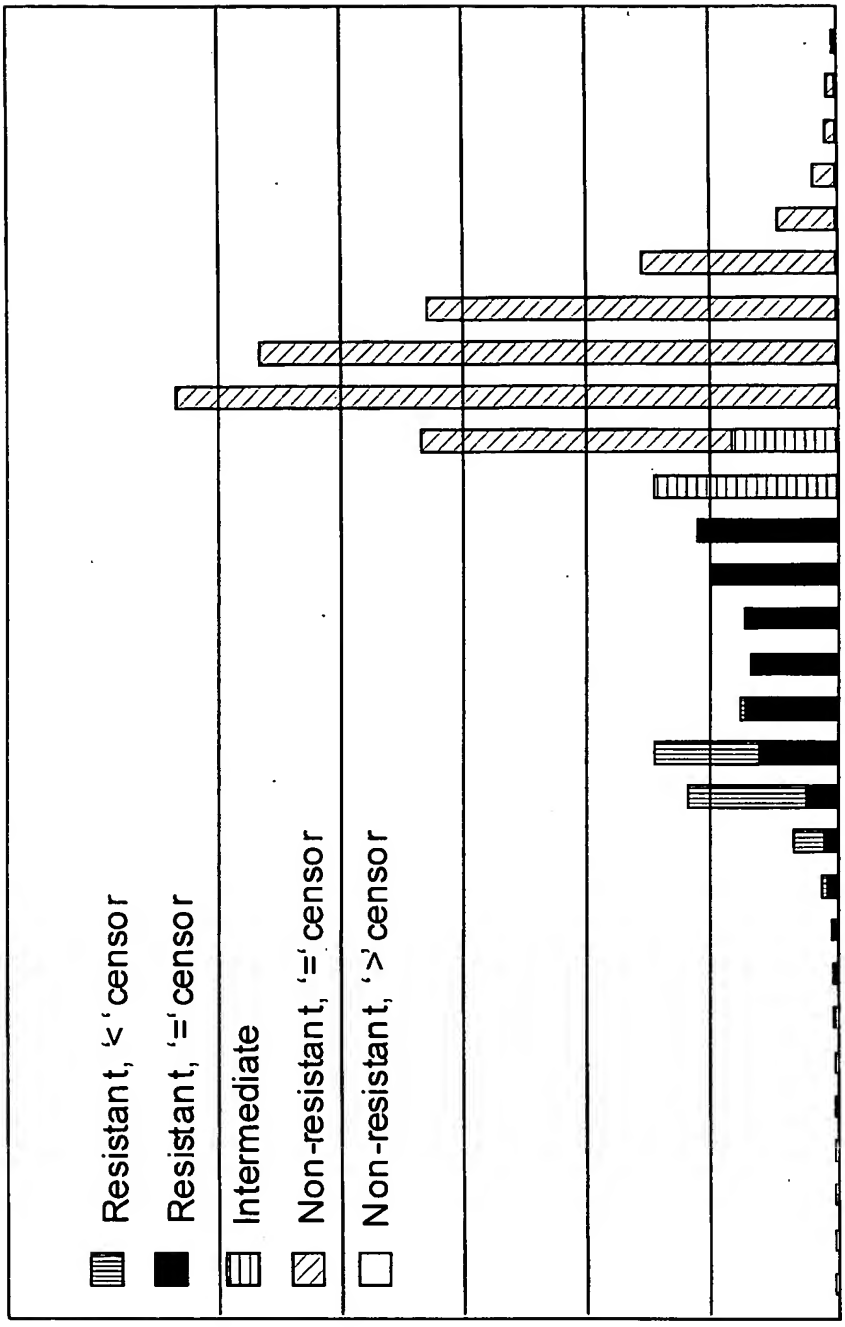

Resistant, < 'censor
Resistant, '=' censor
Intermediate
Non-resistant, '=' censor
Non-resistant, '>' censor
pFR

FIG. 4

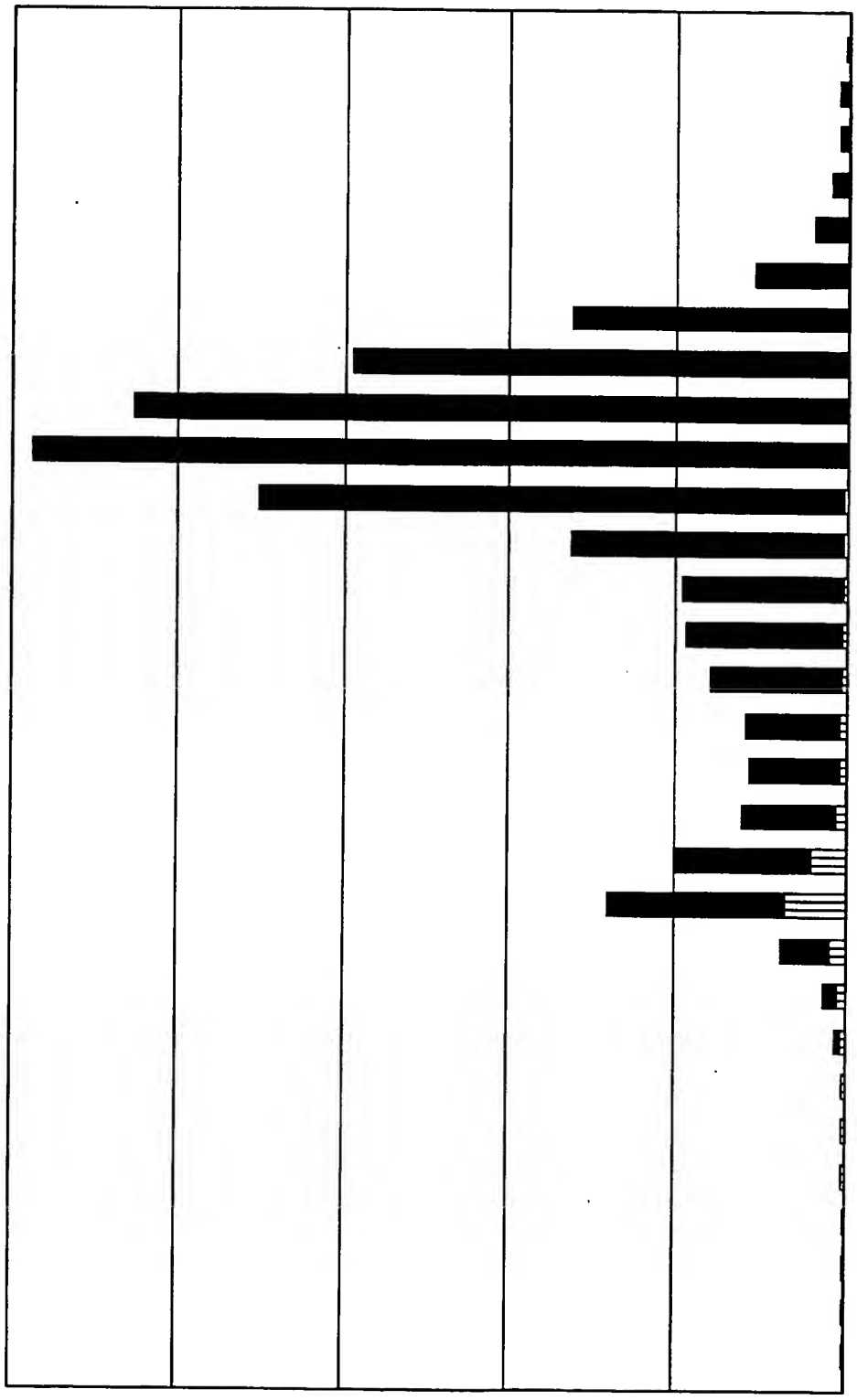
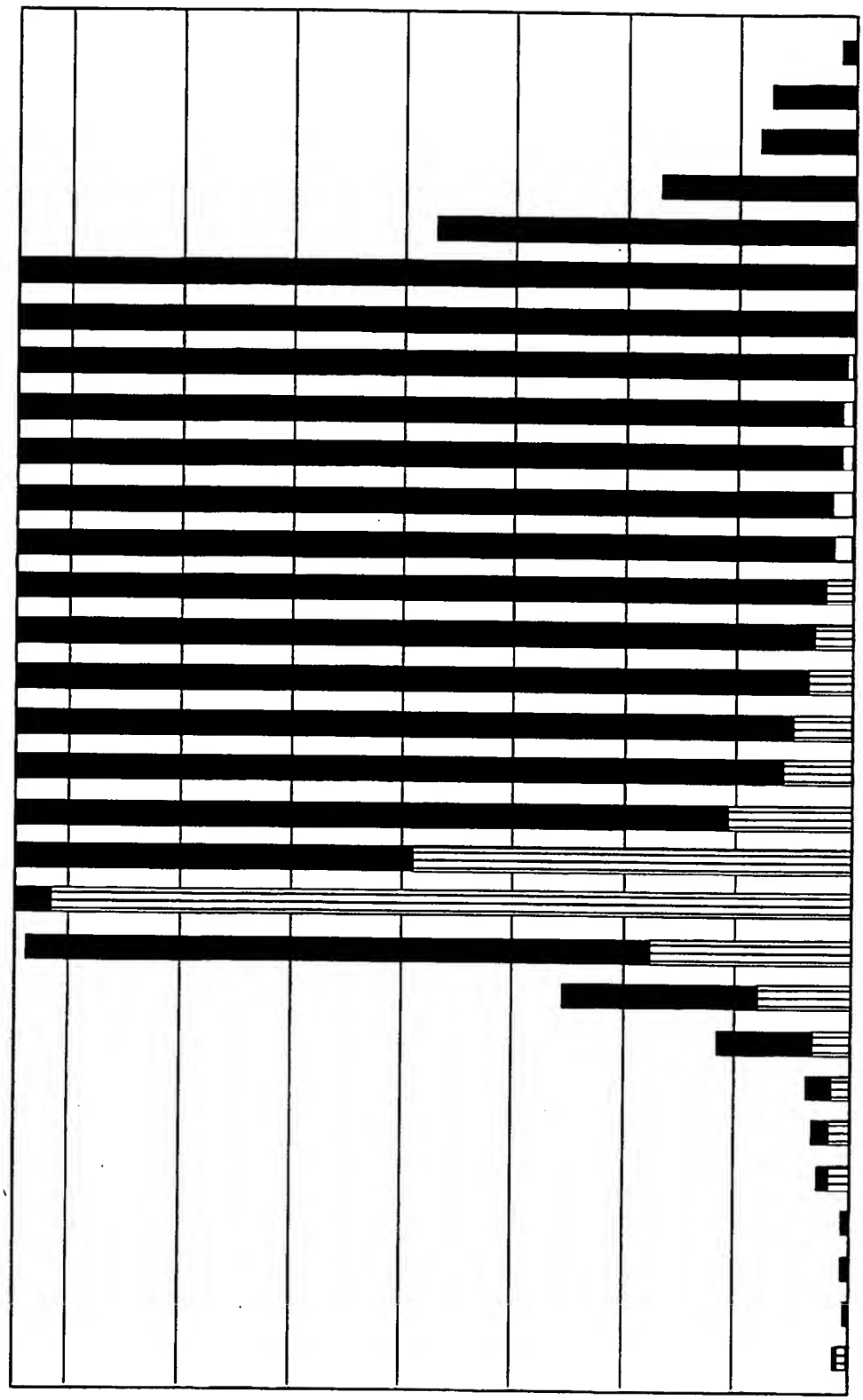


FIG. 5



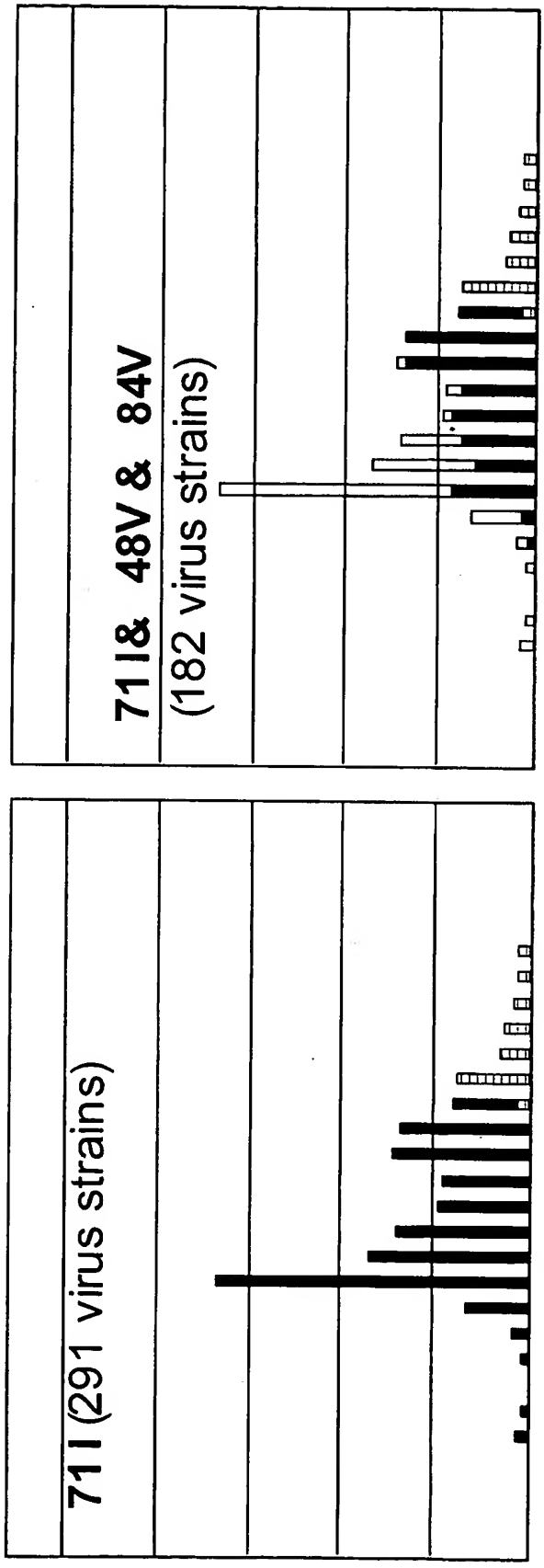


FIG. 6

FIG. 7

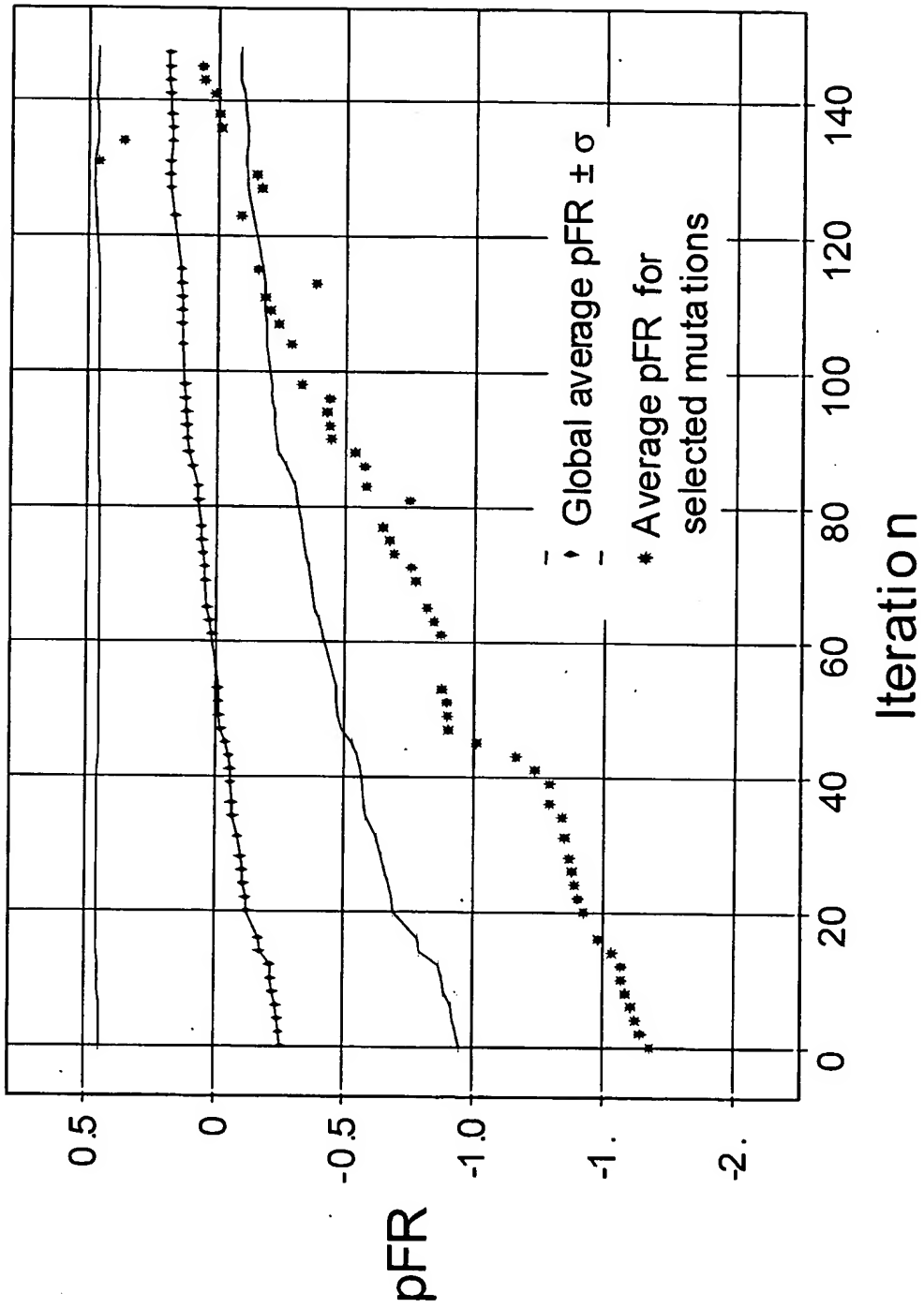

pFR
0.5
0
-0.5
-1.0
-1.
-2.
0
20
40
60
80
100
120
140
Iteration
Global average pFR ± σ
Average pFR for selected mutations

FIG. 8

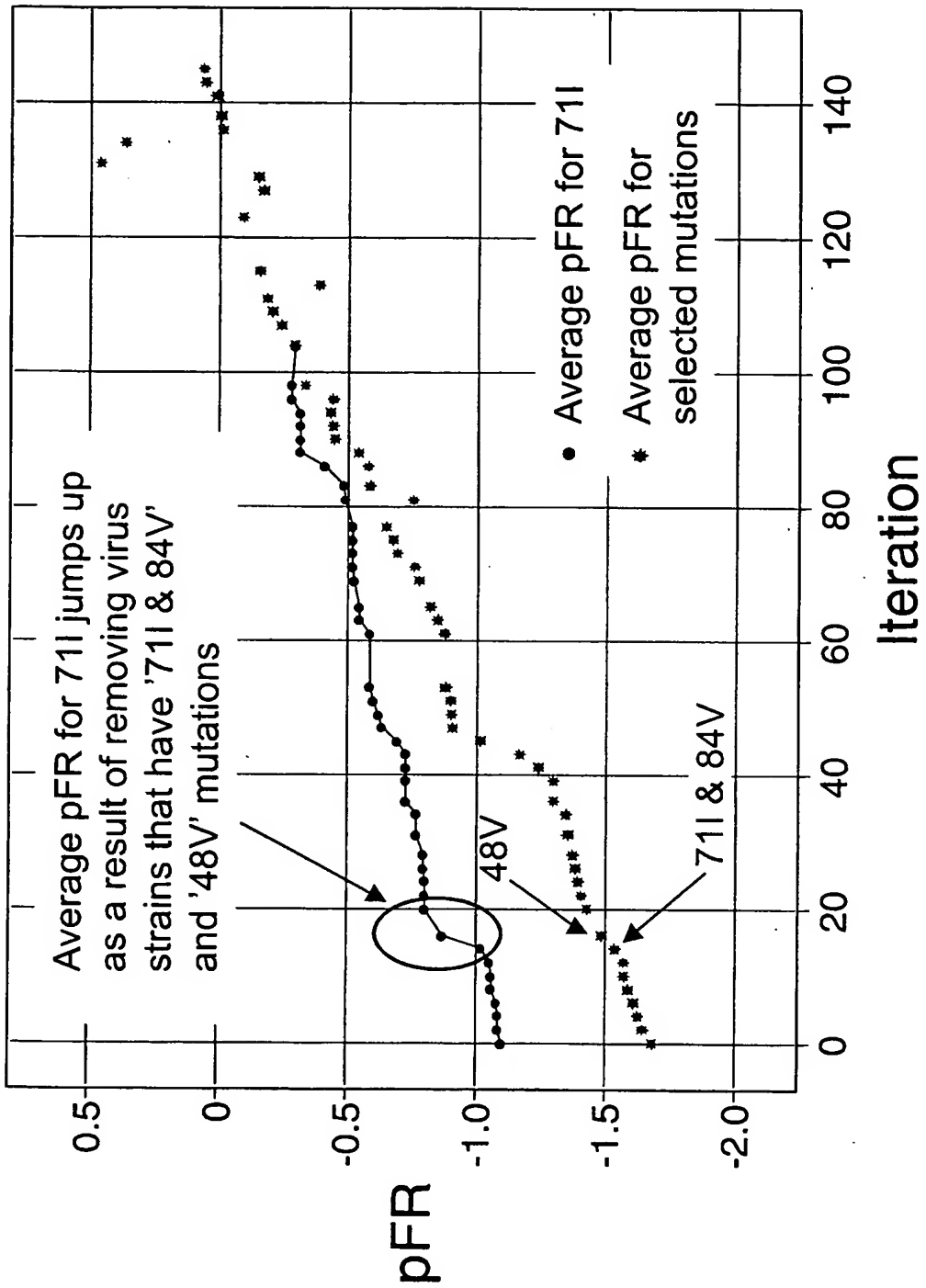

Average pFR for 71I jumps up as a result of removing virus strains that have '71I & 84V' and '48V' mutations
48V
71I & 84V
Average pFR for 71I
Average pFR for selected mutations
pFR
Iteration
0.5
0
-0.5
-1.0
-1.5
-2.0
0
20
40
60
80
100
120
140

FIG. 9

Sample	Position																								
	3	10	12	15	19	20	24	30	32	36	37	41	46	48	54	62	63	71	74	75	77	82	84	93
V1	I								I/V		N		I				P					A		L
V2	I	I		V		R				I	N	K		V	T		T	V	A			A		L
V3	I		S		I			N			N					V	P			I	I			L
V4	I						I				N		I/M		I/V		P	V				T	I/V	

FIG. 10

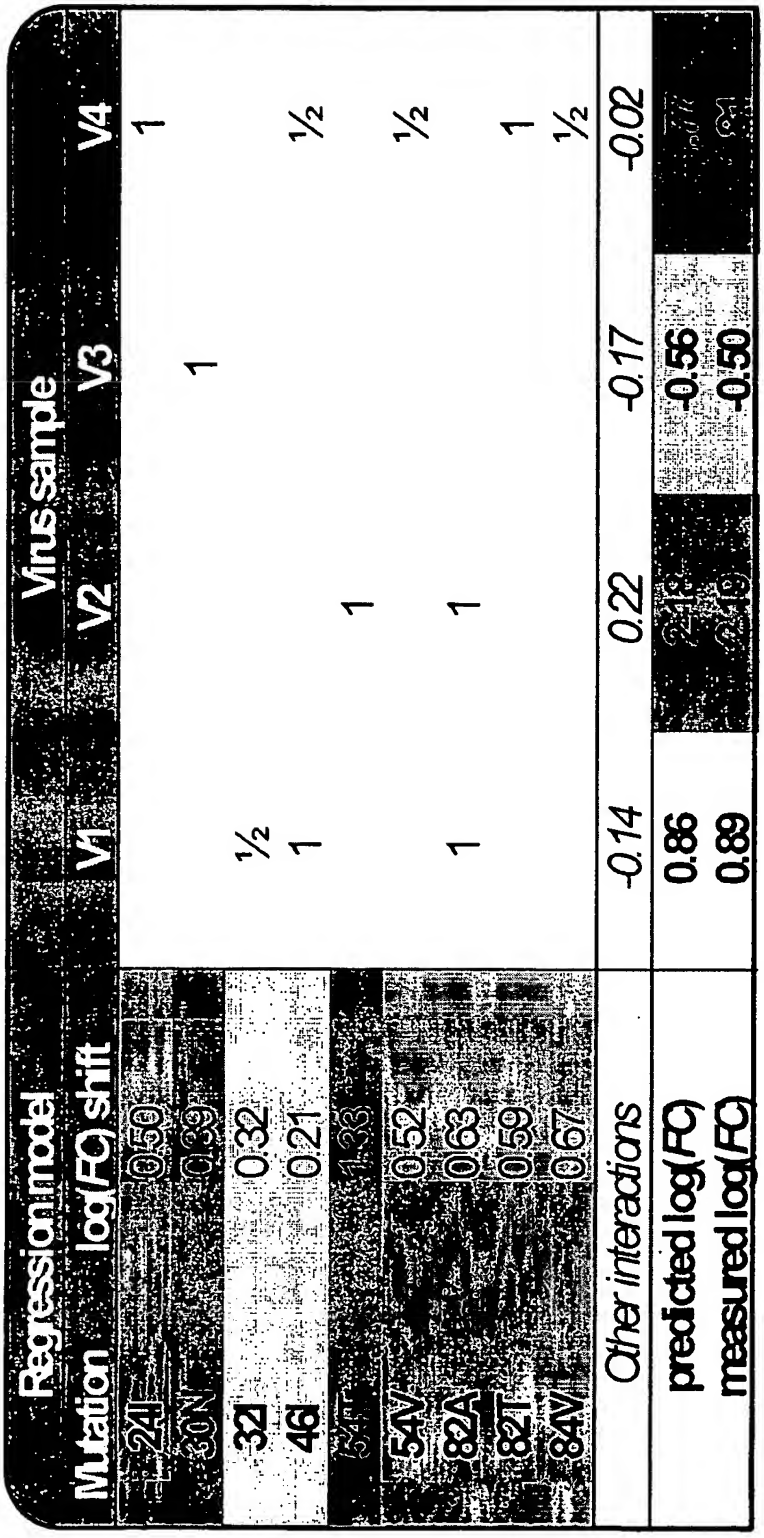

Regression model
Mutation
log(FC) shift
24I 0.50
30N 0.39
32 0.32
46 0.21
54T 1.33
54V 0.52
82A 0.63
82T 0.59
84V 0.67
Other interactions -0.14
predicted log(FC) 0.86
measured log(FC) 0.89
Virus sample
V1
V2
V3
V4
½
1
1
1
1
1
1
½
½
1
½
0.22
-0.17
-0.02
2.18
2.19
-0.56
-0.50

FIG. 11

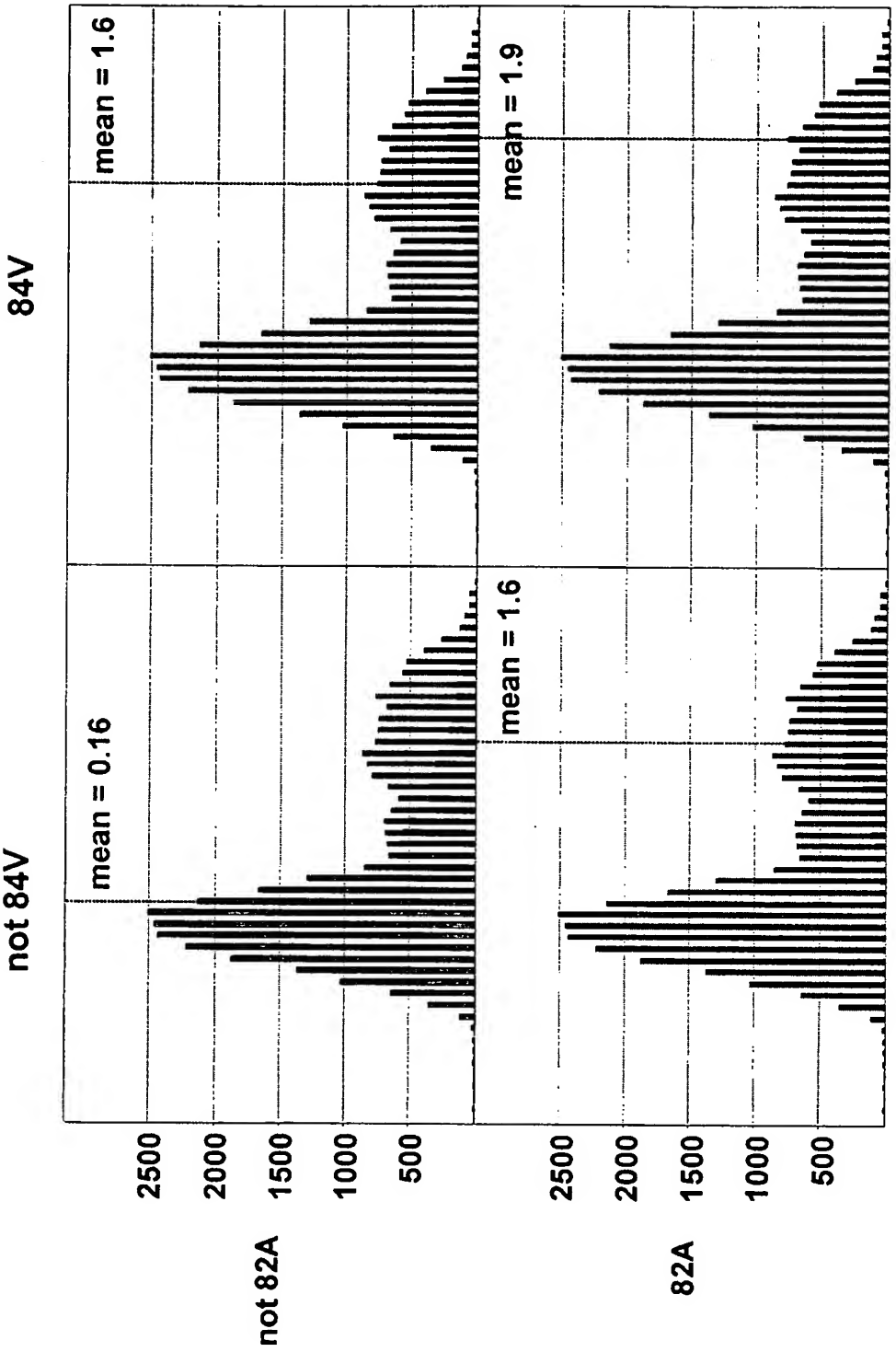

not 84V
84V
mean = 0.16
mean = 1.6
mean = 1.6
mean = 1.9
not 82A
82A
2500
2000
1500
1000
500

FIG. 12

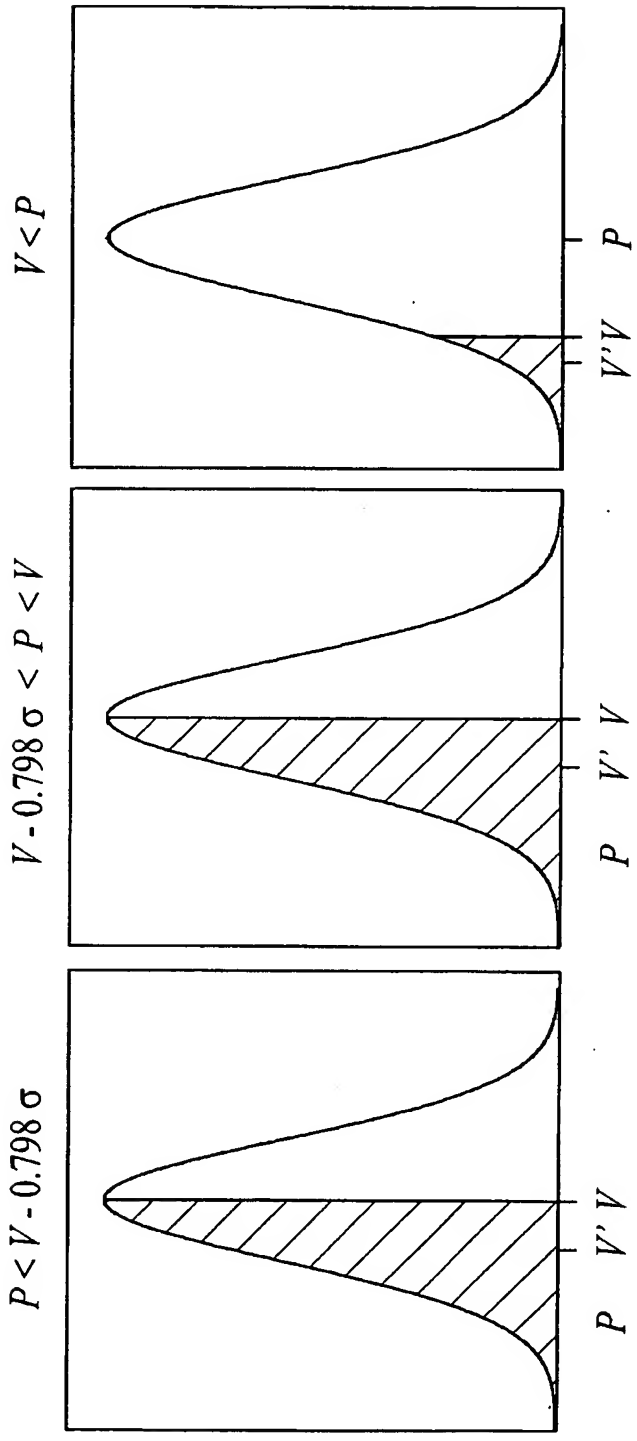


FIG. 13

Mutation	1st order log(FC) shift	Prevalence in dataset
10I*	2nd order terms only	9,707
10R	0.35	106
10V*	0.15	1,269
20M	2nd order terms only	436
20R*	2nd order terms only	2,093
32I	0.32	845
33F*	2nd order terms only	1,074
36I*	2nd order terms only	8,473
46I*	0.21	4,115
46L	0.19	1,745
54L	0.33	367
54V*	0.52	4,553
71T	2nd order terms only	2,611
71V	2nd order terms only	7,261
82A*	0.63	4,886
82F	0.92	290
82T*	0.59	642
82S	1.17	120
84V*	0.67	3,558
90M*	0.38	9,609

FIG. 14

Mutation	log(*FC*) shift	Prevalence in dataset
24I	0.50	1,027
30N	-0.39	1,715
54T	1.33	155
73C	0.45	357
73S	0.38	2,224
73T	0.53	559
82M	0.66	33
84A	1.73	70
84C	0.79	67

FIG. 15

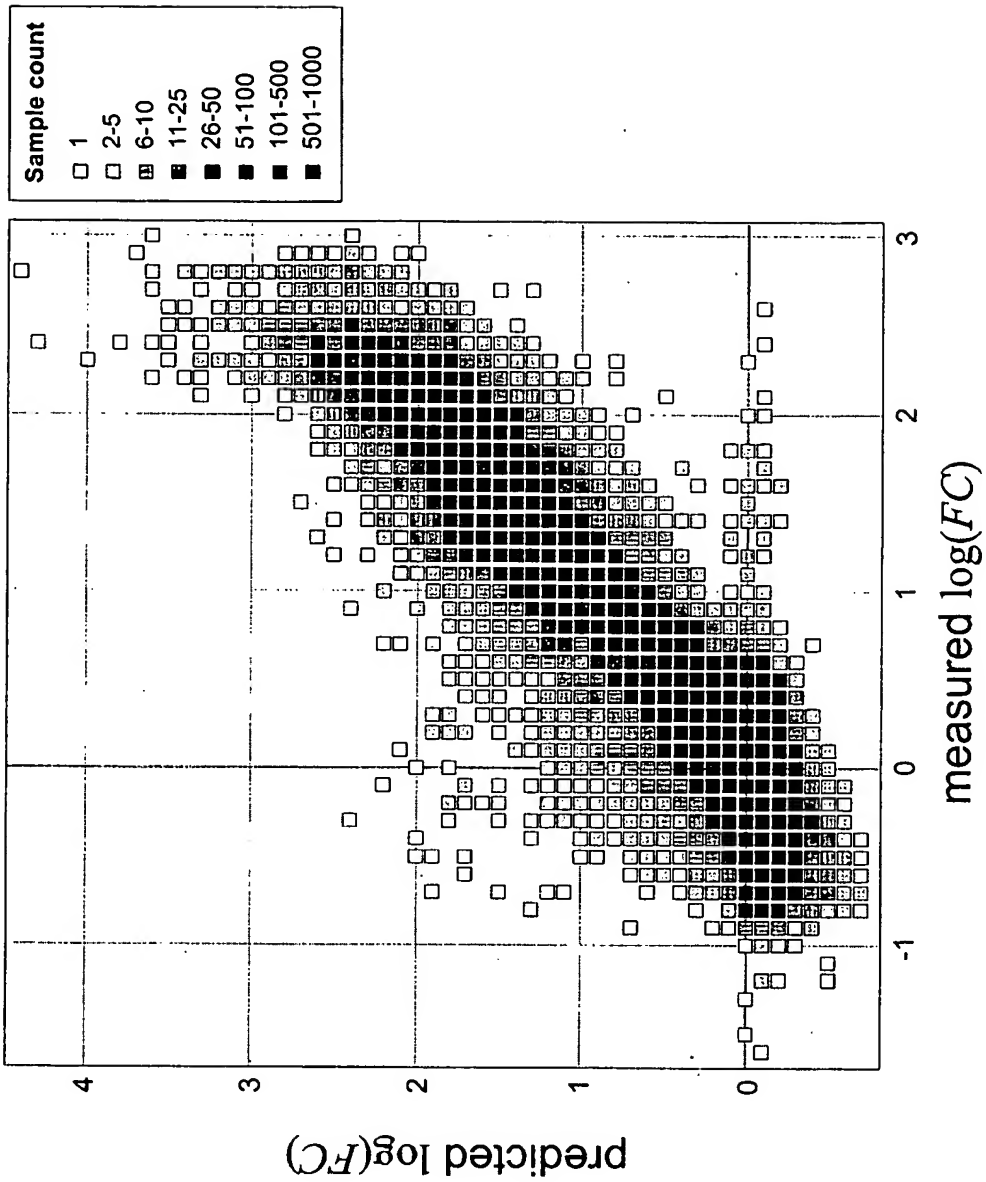

Sample count
1
2-5
6-10
11-25
26-50
51-100
101-500
501-1000
predicted log(FC)
measured log(FC)
0
1
2
3
4
-1

FIG. 16

	Nr. of samples	Resistant fraction (*FC*>3.5)	Leave-one-out prediction error	Sensitivity	Specificity
Decision tree	469	50.1%	10.2%	89.8%	89.7%
Linear model	469	50.1%	6.4%	92.9%	94.4%
Linear model	34,502	38.3%	5.6%	93.0%	95.4%

- **Regression model identifies** 53 single mutations **and** 96 pairs of mutations **as having a positive or negative contribution to RTV susceptibility**
  20 out of 22 **mutations from** IAS list[1] **are confirmed to be significant by regression model**

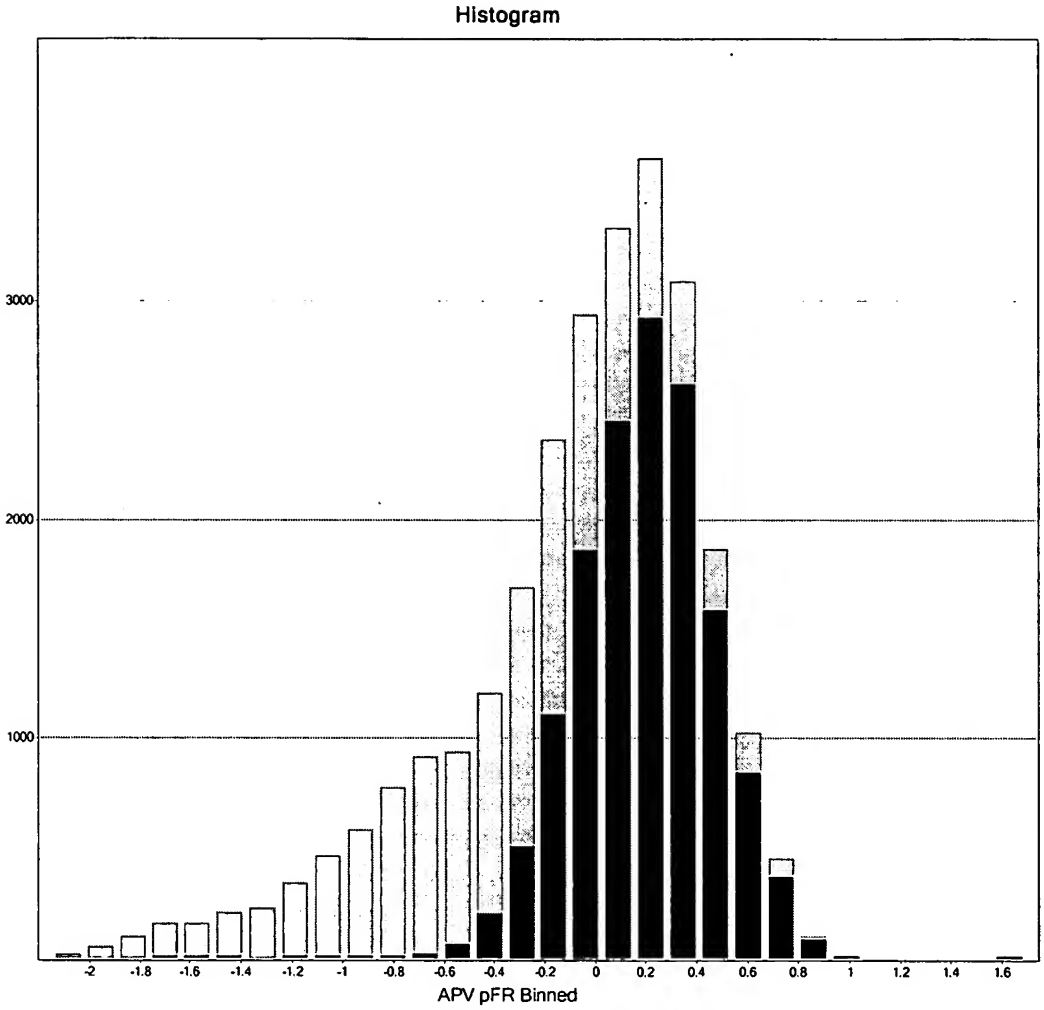

Histogram
3000
2000
1000
-2
-1.8
-1.6
-1.4
-1.2
-1
-0.8
-0.6
-0.4
-0.2
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
APV pFR Binned

FIG. 17

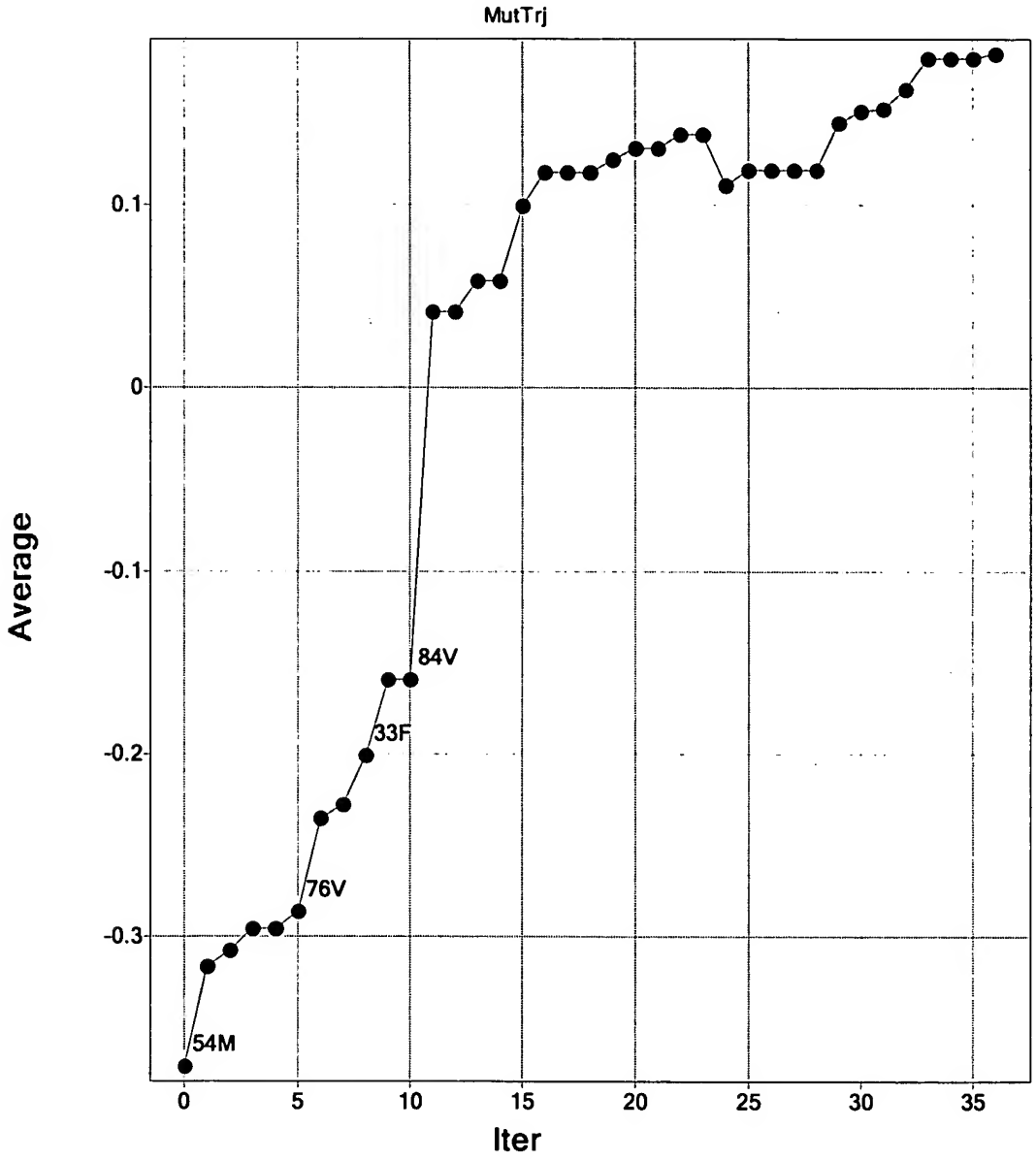

MutTrj
Average
0.1
0
-0.1
-0.2
-0.3
54M
76V
33F
84V
0
5
10
15
20
25
30
35
Iter

FIG. 18

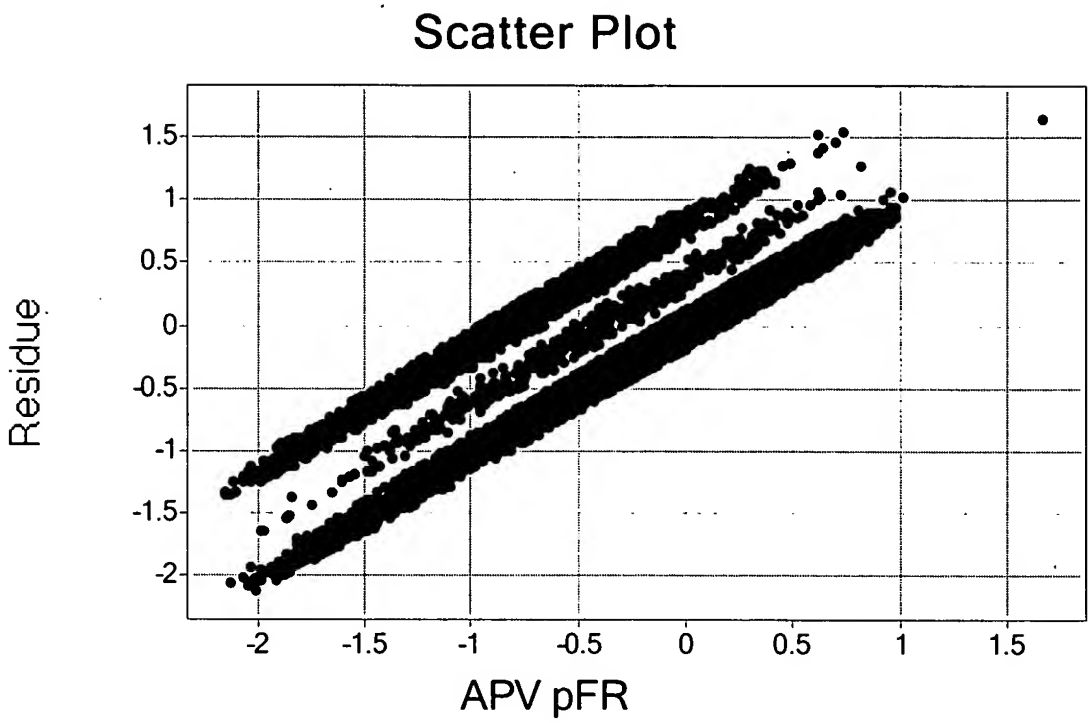


FIG. 19

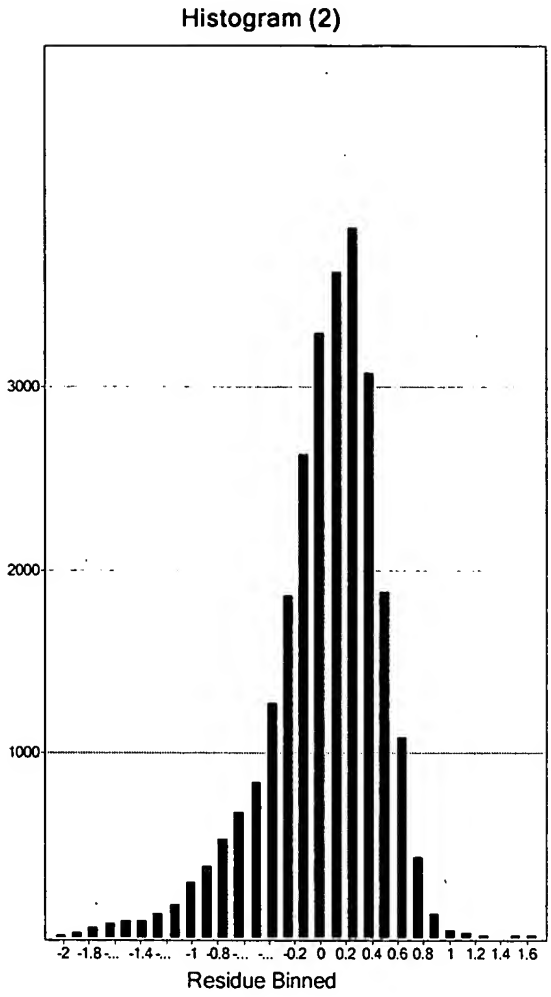

Histogram (2)
3000
2000
1000
-2 -1.8 -... -1.4 -... -1 -0.8 -... -... -0.2 0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
Residue Binned

FIG. 20

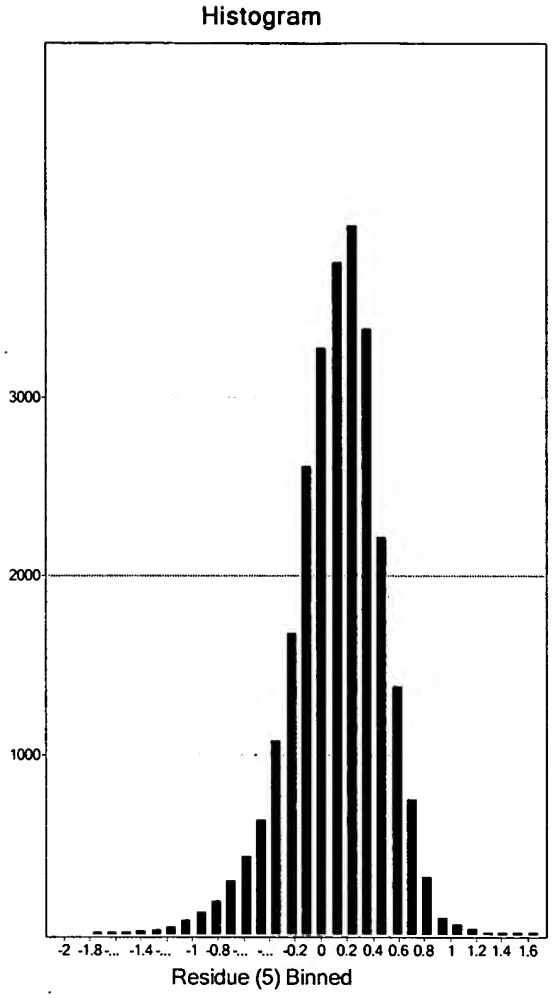

Histogram
3000
2000
1000
-2 -1.8 -... -1.4 -... -1 -0.8 -... -... -0.2 0 0.2 0.4 0.6 0.8 1 1.2 1.4 1.6
Residue (5) Binned

FIG. 21